Family wise separation rates for multiple testing

Magalie Fromont, Matthieu Lerasle, Patricia Reynaud-Bouret

Submitted to the Annals of Statistics

FAMILY WISE SEPARATION RATES FOR MULTIPLE TESTING

BY MAGALIE FROMONT* AND MATTHIEU LERASLE† AND PATRICIA REYNAUD-BOURET†

* *Université Européenne de Bretagne, CNRS, IRMAR, UMR 6625, 35043 Rennes Cedex, France.* † *Université de Nice Sophia-Antipolis, CNRS, LJAD, UMR 7351, 06100 Nice, France.*

Abstract. Starting from a parallel between some minimax adaptive tests of a single null hypothesis, based on aggregation approaches, and some tests of multiple hypotheses, we propose a new second kind error-related evaluation criterion, as the core of an emergent minimax theory for multiple tests.

Aggregation-based tests, proposed for instance by [1], [2], or [6], are justified through their first kind error rate, which is controlled by the prescribed level on the one hand, and through their separation rates over various classes of alternatives, rates that are minimax on the other hand. We show that these tests can be viewed as the first steps of classical step-down multiple testing procedures, and accordingly be evaluated from the multiple testing point of view also, through a control of their Family-Wise Error Rate (FWER). Conversely, many multiple testing procedures, from the historical ones of Bonferroni and Holm, to more recent ones like min-p procedures or randomized procedures such as the ones proposed by [24], can be investigated from the minimax adaptive testing point of view. To this end, we extend the notion of separation rate to the multiple testing field, by defining the *weak Family-Wise Separation Rate* and its stronger counterpart, the *Family-Wise Separation Rate* (FWSR). As for non-parametric tests of a single null hypothesis, we prove that these new concepts allow an accurate analysis of the second kind error of a multiple testing procedure, leading to clear definitions of minimax and minimax adaptive multiple tests.

Some illustrations in a classical Gaussian framework corroborate several expected results under particular conditions on the tested hypotheses, but also lead to more surprising results.

Keywords and phrases : Multiple testing, family-wise error rate, step-down procedure, separation rate, minimax test, adaptive test
AMS 2010 subject classification : 62G10, 62H15, 62C20

1. Introduction.

Following the Neyman-Pearson principle in single null hypotheses testing problems, the main concern in multiple testing problems is

generally to construct procedures controlling a chosen first kind error-related criterion.
Many first kind error-related criteria for multiple tests have been introduced in the statistical literature, generalizing or relaxing the traditional *Family Wise Error Rate* (FWER) defined as the probability of one or more false discoveries (true null hypotheses that are rejected). Thus, the *Per-Family Error Rate* (PFER) suggested by Spjotvoll [28] corresponds to the average number of false discoveries, while the $k-$FWER introduced by Hommel and Hoffman [13] and further studied by Korn et al. [16], Lehmann and Romano [19], Romano and Shaikh [23] or Romano and Wolf [25, 26], is the probability of k or more false discoveries. Like Genovese and Wasserman [10], many of these authors also focused on the *False Discovery Proportion* (FDP), whose expected value is the very popular *False Discovery Rate* (FDR) introduced by Benjamini and Hochberg [3].
Up to now however, very few articles deal with the optimality of multiple tests in terms of second kind error. The articles by Lehmann, Romano, and Shaffer [20], and by Romano, Shaikh and Wolf [22] both give maximin type optimality results, but each with a different notion of maximin optimality. While Lehmann, Romano, and Shaffer [20] consider the minimum probability of rejecting one or more false hypotheses when at least one hypothesis deviates from the truth at a given degree, Romano, Shaikh, and Wolf [22] consider the minimum probability of rejecting at least one hypothesis when the hypotheses are not all true simultaneously.
We propose here new second kind error-related criteria to evaluate multiple procedures whose FWER is controlled by a prescribed level α in $(0, 1)$. These criteria are inspired by the minimax theory for non-parametric tests of a single null hypothesis. The minimax testing theory was historically introduced by Ingster in its series of papers [14] from a purely asymptotic point of view. This asymptotic theory does not seem to be the most adequate to import in multiple testing frameworks, as the asymptotics there should concern the number of tested hypotheses as well as the sample size, leading to consider how the number of hypotheses grows with respect to the sample size. We therefore turned towards the non-asymptotic theory introduced by Baraud in [1], which is based on the notion of uniform separation rate over a class of alternatives. Considering a single null hypothesis H_0 and a class of alternatives $\mathcal{Q}$, the uniform separation rate of a level α test over $\mathcal{Q}$ with prescribed second kind error rate β in $(0, 1)$ is defined as the minimal distance between the underlying distributions in $\mathcal{Q}$ and H_0 which guarantees that the second kind error rate of the test is at most equal to β (a more precise expression is given later on). The test is then said to be minimax over $\mathcal{Q}$ if its uniform

separation rate over $\mathcal{Q}$ achieves the lowest possible value, possibly up to a multiplicative constant. Furthermore, it is said to be minimax adaptive over a collection of classes of alternatives if it is minimax or nearly minimax over every class of the collection.

The literature on minimax and minimax adaptive testing is huge, and provides a now well-known and convenient framework to study the theoretical performance of non-parametric tests of single null hypotheses. Beyond the founding articles by Ingster [14] and Baraud [1], many others are devoted to the computation of minimax separation rates over various classes of alternatives, and the construction of minimax or minimax adaptive tests in many statistical models. For the present concerns, one can cite for instance [29], [2], [15], [6], [8] or [9].

Our purpose here is to provide such a framework in the multiple testing context. Most of minimax adaptive tests of a single null hypothesis H_0 are constructed from the aggregation of a collection of minimax tests for different single null hypotheses, all including H_0. Therefore we first investigate the parallel that can be drawn between such minimax adaptive tests, and some classical single-step or step-down multiple testing procedures. We prove in particular that some of the minimax adaptive tests proposed in [2], [6], [8] or [9] for instance are closely related to the Bonferroni-type single-step multiple testing procedures, while others correspond to the first step of a min-p procedure, as defined in [5]. Conversely, a multiple test may be associated with an aggregated test of a null hypothesis contained in all the tested hypotheses, test that can be studied using the minimax theory. This parallel motivates the definition of the first criterion we introduce here: the *weak Family Wise Separation Rate* denoted by wFWSR, and a stronger second criterion: the (strong) *Family Wise Separation Rate* denoted by FWSR. This second criterion is in fact the key point to lay the foundations of a minimax theory for multiple tests whose FWER is controlled by a prescribed level α. The notion of *minimax Family Wise Separation Rate* presented in this article can thus be used as a new benchmark for the second kind error performance of a multiple test. Beyond the evaluation of a multiple test itself, the *minimax Family Wise Separation Rate* can also be viewed as an indicator of the difficulty or complexity of the considered testing problem. In particular, we exhibit general conditions on the considered hypotheses, which guarantee that the *minimax Family Wise Separation Rate* for multiple tests is lower bounded by the classical minimax Separation Rate for single tests, thus formalizing the intuition that multiple testing is more difficult than single testing. Through an illustration in a simple Gaussian regression framework, we however prove that when these general conditions are not satisfied, the *min-*

imax Family Wise Separation Rate for multiple tests may be much smaller than the classical minimax Separation Rate for single tests, suggesting that multiple testing may be much easier than single testing in some cases. Up to our knowledge, this is a new, counter-intuitive and therefore surprising result that encourages to further develop the minimax theory for multiple tests in future works.
This article is organized as follows. Section 2 contains notations and preliminary results, which may seem to be obvious for the minimax community on the one hand, for the multiple testing community on the other hand. This review is however useful to join the theories from both communities. In Section 3, we investigate the parallel that can be drawn between aggregation-based minimax adaptive tests and some classical multiple testing procedures, leading to the definitions of the wFWSR and the FWSR for multiple testing procedures. We present general results about these notions, as well as a careful study of a simple Gaussian regression framework. The proofs of the main results stated in this section are finally postponed to Section 4.

2. Preliminaries.

Notations. Let X be an observed random variable taking values in a measurable space $(\mathbb{X}, \mathcal{X})$, whose unknown distribution P belongs to a set $\mathcal{P}$ of possible probability distributions on $(\mathbb{X}, \mathcal{X})$.

Following Goeman and Solari [11], an hypothesis H is a subset of $\mathcal{P}$ and H is true under P if P belongs to H, and false under P otherwise.
Given a finite collection $\mathcal{H}$ of such hypotheses, the aim is simultaneously testing H against $\mathcal{P} \setminus H$, for every H in $\mathcal{H}$, which is equivalent to simultaneously testing "H is true under P" against "H is not true under P", or "$P \in H$" against "$P \notin H$", for every H in $\mathcal{H}$.
The set of true hypotheses under P is defined by

$$\mathcal{T}(P) = \{H \in \mathcal{H} \quad / \quad P \in H\} \ ,$$

while the set of false hypotheses under P is

$$\mathcal{F}(P) = \mathcal{H} \setminus \mathcal{T}(P) = \{H \in \mathcal{H} \quad / \quad P \notin H\} \ .$$

A multiple testing procedure or a multiple test may be defined as a statistic given by a collection of rejected hypotheses $\mathcal{R} \subset \mathcal{H}$, only depending on the observed random variable X, whose goal is to infer the set $\mathcal{F}(P)$ of false hypotheses under P in $\mathcal{H}$. In the following, $\cap\mathcal{H}$ is an abbreviation for $\cap_{H\in\mathcal{H}} H$.

Most of the tests presented here are based on test statistics, for which we need the following formalism. For any real valued statistic T, it is classical to consider its cumulative distribution function (c.d.f.) F, which is càdlàg, and its corresponding generalized inverse function or quantile function F^{-1} which is càglàd. However in the sequel, we focus on the càglàd c.d.f. of T, denoted by F_- and defined by

$$\forall t \in \mathbb{R}, \quad F_-(t) = P(T < t) \ .$$

Its generalized inverse function F_-^{-1} is then a càdlàg function defined by

$$\forall u \in (0,1), \quad F_-^{-1}(u) = \sup\{t,\ F_-(t) \leq u\} \ .$$

2.1. *Tests of a single null hypothesis and p-values.* A test of a single null hypothesis H_0 is usually formalized as a statistic taking values in $\{0,1\}$, whose value 1 amounts to rejecting H_0.
There are two classical ways of defining such a test, either by giving a test statistic and the corresponding critical values, or by giving a p-value. The following preliminary result allows us to precisely go back and forth between the "multiple tests" literature used to p-values, and the "aggregated tests" literature, exclusively using test statistics and critical values. Notice that a part of the statements of this result can be proved thanks to Lemma 1.1. in [19]. We however give a comprehensive and self-contained proof in Section 4.

LEMMA 1. *Let T be a real-valued test statistic of a single null hypothesis H_0 whose distribution does not depend on P provided that P belongs to H_0. Denote by F and F_- the (càdlàg) c.d.f. and the càglàd c.d.f. of this distribution under H_0, and by F^{-1} and F_-^{-1} their respective generalized inverse functions as defined above. Let $p(T) = 1 - F_-(T)$, and for any fixed level α in $(0,1)$,*

$$\phi = \mathbb{1}_{\{T > F^{-1}(1-\alpha)\}},\ \phi_- = \mathbb{1}_{\left\{T > F_-^{-1}(1-\alpha)\right\}},\ \text{and } \phi_p = \mathbb{1}_{\{p(T) \leq \alpha\}} \ .$$

Then all those three tests are of level α and their associated p-value (i.e. the limit level α at which they pass from acceptance to rejection) is $p(T)$, which satisfies for all P in H_0,

$$P(p(T) \leq \alpha) \leq \alpha \ .$$

Moreover,

$$T > F_-^{-1}(1-\alpha) \Leftrightarrow p(T) < \alpha \ , \tag{1}$$

and

$$\left\{ \begin{array}{lll} \phi_- & \leq & \phi \\ \phi_- & \leq & \phi_p \end{array} \right. . \tag{2}$$

Most of the time, c.d.f. are continuous and in this case the three tests ϕ, ϕ_- and ϕ_p are almost surely equal. However when atoms are present in the distribution, the most powerful one is the test ϕ_p based on the p-value, which is not completely equivalent to the more classical test ϕ based on the test statistic. Note that this lemma also applies if F is not the c.d.f. but a conditional c.d.f. This can be especially useful when bootstrap or permutation procedures are used since, in this case, the considered distributions are naturally non continuous (see [24–26] for instance).
Authors using p-values generally consider the test ϕ_p, while authors used to test statistics and critical values generally consider the test ϕ. These two tests are not always almost surely equal. In order to more conveniently go back and forth between p-values on the one hand, test statistics and critical values on the other hand, regarding the equivalence stated in (1), we focus all along the paper on tests in the form of:

$$\phi_- = \mathbb{1}_{\left\{T > F_-^{-1}(1-\alpha)\right\}} = \mathbb{1}_{\{p(T) < \alpha\}} .$$

In particular, when we refer in the sequel to well-known procedures such as Bonferonni or Holm's ones, we in fact refer to the versions of these procedures written in the form of ϕ_- above.
Note that the test ϕ_p, generally considered by authors, can also be expressed using test statistics and critical values (see Corollary 10 in Section 4), but at the price of a much more intricate formula.

2.2. *Multiple tests and the Family-Wise Error Rate.* The weak Family-Wise Error Rate of a multiple test $\mathcal{R}$ is denoted by $w\mathrm{FWER}(\mathcal{R})$ and is defined by:

$$w\mathrm{FWER}(\mathcal{R}) = \sup_{P/\mathcal{T}(P)=\mathcal{H}} P\left(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset\right) . \tag{3}$$

Controlling the wFWER is generally too weak in applications, as some of the hypotheses in $\mathcal{H}$ may actually be false under P. A control of the probability $P\left(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset\right)$, for any possible P, is therefore more appropriate. This leads to the following definition of the (strong) Family-Wise Error Rate of $\mathcal{R}$, denoted by $\mathrm{FWER}(\mathcal{R})$:

$$\mathrm{FWER}(\mathcal{R}) = \sup_{P \in \mathcal{P}} P\left(\mathcal{R} \cap \mathcal{T}(P) \neq \emptyset\right) . \tag{4}$$

Given a prescribed level α in $(0,1)$, the main concern then becomes to construct a multiple test $\mathcal{R}$ such that

$$\text{FWER}\,(\mathcal{R}) \leq \alpha \ , \tag{5}$$

which obviously also implies that $w\text{FWER}\,(\mathcal{R}) \leq \alpha$.
A large number of multiple tests satisfying (5) have been constructed, among them the historical procedures of Bonferroni and Holm [12, 27], and the more recent min-p type procedures (see [5] for instance). Many of these procedures can be described through the general sequential rejection scheme proposed by Goeman and Solari [11], which consists in iteratively rejecting hypotheses through an application $\mathcal{N}$ from the set of all subsets of $\mathcal{H}$ to itself, as follows.

$$\begin{cases} 1. \quad \text{Start with } \mathcal{R}_0 = \emptyset \ . \\ 2. \quad \text{For any } n \geq 0, \text{ build } \mathcal{R}_{n+1} = \mathcal{R}_n \cup \mathcal{N}(\mathcal{R}_n) \ . \\ 3. \quad \text{Define } \mathcal{R} = \lim_{n\to\infty} \mathcal{R}_n \ . \end{cases} \tag{6}$$

Notice that the sequence $(\mathcal{R}_n)_{n\geq 0}$ is always convergent in the present framework since $\mathcal{H}$ is assumed to be finite.
For any prescribed α in $(0,1)$, Goeman and Solari [11, Theorem 1]) proved that sequential rejective procedures satisfy (5), as soon as the two conditions below, named (**Monotonicity**) and (**Single − Step**), are true:

$$(\textbf{Monotonicity}) \qquad \forall \mathcal{S} \subset \mathcal{S}' \subset \mathcal{H},\ \mathcal{N}(\mathcal{S}) \subset \mathcal{S}' \cup \mathcal{N}(\mathcal{S}') \ ,$$

$$(\textbf{Single}-\textbf{Step}) \qquad \forall P \in \mathcal{P},\ P\left(\mathcal{N}(\mathcal{F}(P)) \subset \mathcal{F}(P)\right) \geq 1-\alpha \ .$$

Let us focus on a generic example, the min-p procedure, assuming that a set $\mathcal{H}$ of hypotheses and their corresponding p-values p_H, for H in $\mathcal{H}$, are given, such that for all P in H,

$$\forall u \in (0,1), \quad P(p_H \leq u) \leq u \ .$$

For any subset $\mathcal{G}$ of $\mathcal{H}$, and any α in $(0,1)$, let $q_{mp,\mathcal{G},\alpha}$ be a non increasing function of $\mathcal{G}$ such that

$$\forall P \in \cap\mathcal{G},\ P\left(\min_{H\in\mathcal{G}} p_H < q_{mp,\mathcal{G},\alpha}\right) \leq \alpha \ .$$

Then the min-p procedure is defined as a sequential rejective procedure with the application $\mathcal{N}$ equal to

$$\mathcal{N}_{mp} : \mathcal{S} \mapsto \left\{ H \in \mathcal{H} \setminus \mathcal{S} \ \ / \ \ p_H < q_{mp,\mathcal{H}\setminus\mathcal{S},\alpha} \right\} \ .$$

As it satisfies (**Monotonicity**) and (**Single − Step**), by [11, Theorem 1]), the min-p procedure has a FWER controlled by α.
It is always possible to use $q_{mp,\mathcal{G},\alpha} = \alpha/\#\mathcal{G}$, where $\#\mathcal{G}$ denotes the cardinal of $\mathcal{G}$. The obtained multiple test is due to Holm [12], so we denote it by $\mathcal{R}_{Holm}$ and the corresponding application by $\mathcal{N}_{Holm}$.
The first step of this procedure corresponds to the well-known Bonferroni procedure and is denoted by $\mathcal{R}_{Bonf} := \mathcal{N}_{Holm}(\emptyset)$.
A more precise choice can be done as follows. If the distribution of $\min_{H\in\mathcal{G}} p_H$ (with càglàd c.d.f. $F_{\mathcal{G}}$ and càglàd c.d.f. $F_{\mathcal{G},-}$) does not depend on P in $\cap\mathcal{G}$ and is known, one can take $q_{mp,\mathcal{G},\alpha} = F^{-1}_{\mathcal{G},-}(\alpha)$. The resulting rejection set is then denoted by $\mathcal{R}_{mp}$. Note that this multiple testing procedure is less conservative than $\mathcal{R}_{Holm}$, i.e.

$$\mathcal{R}_{Holm} \subset \mathcal{R}_{mp} \quad .$$

If $F_{\mathcal{G}}$ is unknown, the quantiles may be replaced by random quantiles, depending on X, based on permutation or bootstrap approaches [24–26], at the possible price of an asymptotic control of the FWER instead of an exact control.
The min-p procedures may also be extended to weighted min-p procedures by defining

$$\mathcal{N}_{wmp} : \mathcal{S} \mapsto \left\{ H \in \mathcal{H} \setminus \mathcal{S} \;\;/\;\; p_H < w_H \, q_{wmp,\mathcal{H}\setminus\mathcal{S},\alpha} \right\} \quad ,$$

where $(w_H)_{H\in\mathcal{H}}$ is a family of positive weights satisfying $\sum_{H\in\mathcal{H}} w_H = 1$, and where $q_{wmp,\mathcal{G},\alpha}$ satisfies for any α in $(0, 1)$,

$$\forall P \in \cap\mathcal{G},\;\; P\left(\min_{H\in\mathcal{G}} w_H^{-1} p_H < q_{wmp,\mathcal{G},\alpha}\right) \leq \alpha \quad .$$

When the distribution of $\min_{H\in\mathcal{G}} w_H^{-1} p_H$ (with càglàd c.d.f. $F_{w,\mathcal{G},-}$) does not depend on P in $\cap\mathcal{G}$ and is known, one can take $q_{wmp,\mathcal{G},\alpha} = F^{-1}_{w,\mathcal{G},-}(\alpha)$, which defines rejection sets denoted by $\mathcal{R}_{wmp}$. Note that these last procedures are very close to the balanced procedure of Romano and Wolf [26].

2.3. *Aggregated tests and the First Kind Error Rate.* Considering the problem of testing a single null hypothesis H_0 against the alternative $\mathcal{P} \setminus H_0$, we sketch a general methodology for the construction of aggregated tests, and then focus on a classical example.
The idea of aggregated tests comes from the minimax adaptivity theory. Indeed, the construction of minimax adaptive tests, as defined by Spokoiny [29], often consists in the aggregation of a finite collection of initial minimax (non-adaptive) individual tests.

In general, a finite collection of hypotheses $\mathcal{H}$ is chosen such that $H_0 \subset \cap\mathcal{H}$, and so that the final aggregated test achieves some expected minimax adaptivity properties. For each hypothesis H in the collection $\mathcal{H}$, an individual test ϕ_H of the null hypothesis H against the alternative $\mathcal{P}\setminus H$ is constructed, that is a statistic with values in $\{0,1\}$, whose value 1 amounts to rejecting H. The collection of tests is denoted $\Phi_{\mathcal{H}} = \{\phi_H, H \in \mathcal{H}\}$. Then, the corresponding aggregated test $\bar{\Phi}_{\mathcal{H}}$ consists in rejecting H_0 if at least one of the H's is rejected with ϕ_H, that is

$$\bar{\Phi}_{\mathcal{H}} = \sup_{H\in\mathcal{H}} \phi_H \ . \tag{7}$$

Note that in the original works, the ϕ_H's are not presented as individual tests of H against $\mathcal{P}\setminus H$, but as some of numerous tests of the original single null hypothesis H_0 against the alternative $\mathcal{P} \setminus H_0$.

Many frameworks have been studied, among them of course Gaussian regression frameworks (see [29], [2], [18] for instance), density or Poisson processes frameworks (see [15], [6], [8]), or more complex ones corresponding to two-sample type problems (see [7], [9], [4]). We focus here on the most simple Gaussian regression framework considered in [1], to illustrate things as clearly as possible.

A Gaussian regression framework. The observed random variable is a random vector $X = (X_1, \ldots, X_n)'$ whose distribution $P = P_{\mathbf{f}}$ is an n-dimensional Gaussian distribution with mean $\mathbf{f}$, and covariance matrix $\sigma^2 I_n$ ($n \geq 1$). The mean $\mathbf{f} = (f_1, ..., f_n)'$ is unknown, while $\sigma^2 > 0$ is assumed to be known.

We consider the problem of testing the single null hypothesis $H_0 = \{P_0\}$ against the alternative $\mathcal{P} \setminus \{P_0\}$, with $\mathcal{P} = \{P_{\mathbf{f}},\ f \in \mathbb{R}^n\}$, that is testing "$\mathbf{f} = 0$" against "$\mathbf{f} \neq 0$".

From a fixed collection $\mathcal{S}$ of vectorial subspaces S of $\mathbb{R}^n$, a collection of tests ϕ_S of H_0 against $\mathcal{P} \setminus H_0$ is constructed, where ϕ_S equals 1 when the norm $\|\Pi_S X\|$ of the orthogonal projection $\Pi_S(X)$ of X onto S (w.r.t. the Euclidean distance) takes large values. Considering the individual hypothesis

$$H_S = \{P_{\mathbf{f}}\ /\ \Pi_S\mathbf{f} = 0\} = \left\{P_{\mathbf{f}}\ /\ \mathbf{f} \in S^{\perp}\right\} \ ,$$

ϕ_S may also be viewed as an individual test of H_S against $\mathcal{P} \setminus H_S$, and can thus be denoted by ϕ_{H_S}.

For $\mathcal{H} = \{H_S,\ S \in \mathcal{S}\}$, the aggregated test of the null hypothesis $H_0 = \{P_0\}$ against the alternative $\mathcal{P} \setminus \{P_0\}$, based on the collection of individual tests $\Phi_{\mathcal{H}} = \{\phi_{H_S},\ S \in \mathcal{S}\}$ is then defined as in (7) by

$$\bar{\Phi}_{\mathcal{H}} = \sup_{H_S\in\mathcal{H}} \phi_{H_S} \ .$$

Notice that if the collection $\mathcal{S}$ is not rich enough, then $H_0 \subsetneq \cap\mathcal{H}$.

First kind error rate of aggregated tests. The first kind error rate of an aggregated test $\bar{\Phi}_{\mathcal{H}}$ of the single null hypothesis H_0 is defined as usual by

$$\mathrm{ER}\left(\bar{\Phi}_{\mathcal{H}}, H_0\right) = \sup_{P\in H_0} P\left(\bar{\Phi}_{\mathcal{H}} = 1\right) = \sup_{P\in H_0} P\left(\sup_{H\in\mathcal{H}} \phi_H = 1\right) \ .$$

This criterion should be controlled by a prescribed level α in $(0,1)$. For any hypothesis H of the collection $\mathcal{H}$, the individual test ϕ_H is usually defined from a test statistic T_H, whose distribution does not depend on P provided that P belongs to H_0. Following Lemma 1 and respectively denoting by $F_{H,-}$ and $F_{H,-}^{-1}$ the càglàd c.d.f. and càdlàg quantile function of this distribution under H_0, ϕ_H is then defined as $\mathbb{1}_{\left\{T_H > F_{H,-}^{-1}(1-u_{H,\alpha})\right\}}$, where $u_{H,\alpha}$ is chosen so that the aggregated test is actually of level α, that is

$$\mathrm{ER}\left(\bar{\Phi}_{\mathcal{H}}, H_0\right) \leq \alpha \ .$$

The most obvious choice for $u_{H,\alpha}$ is a Bonferroni-type choice $u_{H,\alpha} = \alpha/N$, where $N = \#\mathcal{H}$ is the number of hypotheses in $\mathcal{H}$. This leads to the Bonferroni-type aggregated test $\bar{\Phi}_{\mathcal{H}}^{Bonf}$ based on the collection

$$\Phi_{\mathcal{H}}^{Bonf} = \left\{\phi_H^{Bonf} = \mathbb{1}_{\left\{T_H > F_{H,-}^{-1}(1-\alpha/N)\right\}}, H \in \mathcal{H}\right\} \ .$$

A weighted Bonferroni-type choice $u_{H,\alpha} = w_H\alpha$ is also proposed in [8] and [9], where $(w_H)_{H\in\mathcal{H}}$ is a family of positive weights such that $\sum_{H\in\mathcal{H}} w_H \leq 1$. This leads to the weighted Bonferroni-type aggregated test $\bar{\Phi}_{\mathcal{H}}^{wBonf}$ based on the collection

$$\Phi_{\mathcal{H}}^{wBonf} = \left\{\phi_H^{wBonf} = \mathbb{1}_{\left\{T_H > F_{H,-}^{-1}(1-w_H\alpha)\right\}}, H \in \mathcal{H}\right\} \ .$$

A less conservative choice in practice and still guaranteeing a level α is proposed by Baraud, Huet and Laurent [2]. It consists in taking $u_{H,\alpha} = w_H u_\alpha$, where

$$u_\alpha = \sup\left\{u \ \Big/ \ \sup_{P\in H_0} P\left(\exists H \in \mathcal{H} \ / \ T_H > F_{H,-}^{-1}(1-w_H u)\right) \leq \alpha\right\} \ .$$

This leads, when $w_H = 1/N$, to the aggregated test $\bar{\Phi}_{\mathcal{H}}^{BHL}$ based on the collection

$$\Phi_{\mathcal{H}}^{BHL} = \left\{\phi_H^{BHL} = \mathbb{1}_{\left\{T_H > F_{H,-}^{-1}(1-u_\alpha/N)\right\}}, H \in \mathcal{H}\right\} \ ,$$

or, in the general case, to the aggregated test $\bar{\Phi}_{\mathcal{H}}^{wBHL}$ based on the collection

$$\Phi_{\mathcal{H}}^{wBHL} = \left\{\phi_H^{wBHL} = \mathbb{1}_{\left\{T_H > F_{H,-}^{-1}(1-w_H u_\alpha)\right\}}, H \in \mathcal{H}\right\} \ .$$

3. Main results. In this section, we first study the main correspondences between both theories: multiple tests and aggregated tests. To do so, we always assume that a finite collection of hypotheses $\mathcal{H}$ and a single null hypothesis H_0 such that $H_0 \subset \cap\mathcal{H}$ are given.
From any collection $\Phi_{\mathcal{H}} = \{\phi_H, H \in \mathcal{H}\}$ of tests ϕ_H of the single hypothesis H, a multiple test of $\mathcal{H}$ is constructed as

$$\mathcal{R}\left(\Phi_{\mathcal{H}}\right) = \left\{ H \in \mathcal{H} \;/\; \phi_H = 1 \right\} \ .$$

Conversely, from any multiple test $\mathcal{R}$ of $\mathcal{H}$, we construct

$$\bar{\Phi}(\mathcal{R}) = \mathbb{1}_{\{\mathcal{R} \neq \emptyset\}} \ ,$$

which can be seen as a test of the single null hypothesis H_0.

3.1. *First kind error and first identifications.* First notice that the weak Family Wise Error Rate of $\mathcal{R}\left(\Phi_{\mathcal{H}}\right)$ is equal to

$$\begin{aligned} w\text{FWER}\left(\mathcal{R}\left(\Phi_{\mathcal{H}}\right)\right) &= \sup_{P \in \cap\mathcal{H}} P\left(\mathcal{R}\left(\Phi_{\mathcal{H}}\right) \neq \emptyset\right) \\ &= \sup_{P \in \cap\mathcal{H}} P\left(\bar{\Phi}_{\mathcal{H}} = 1\right) \\ &= \text{ER}\left(\bar{\Phi}_{\mathcal{H}}, \cap\mathcal{H}\right) \ . \end{aligned}$$

Since $H_0 \subset \cap\mathcal{H}$,

$$w\text{FWER}\left(\mathcal{R}\left(\Phi_{\mathcal{H}}\right)\right) \geq \text{ER}\left(\bar{\Phi}_{\mathcal{H}}, H_0\right) \ . \tag{8}$$

Except in the case where H_0 exactly equals $\cap\mathcal{H}$, controlling $w\text{FWER}\left(\mathcal{R}\left(\Phi_{\mathcal{H}}\right)\right)$ is thus more difficult than controlling $\text{ER}\left(\bar{\Phi}_{\mathcal{H}}, H_0\right)$.

Next, assume that for every H in $\mathcal{H}$, a test statistic T_H, whose distribution does not depend on P provided that P belongs to H, is given and denote by p_H its corresponding p-value, as defined by Lemma 1.

PROPOSITION 2. *With the notations of Sections 2.2 and 2.3, the following identifications hold:*

$$\mathcal{R}(\Phi_{\mathcal{H}}^{Bonf}) = \mathcal{R}_{Bonf} \ ,$$

and

$$\bar{\Phi}_{\mathcal{H}}^{Bonf} = \bar{\Phi}(\mathcal{R}_{Bonf}) = \bar{\Phi}(\mathcal{R}_{Holm}) \ .$$

If additionally the distribution of $\min_{H \in \mathcal{H}} w_H^{-1} p_H$ *does not depend on* P *provided that* P *belongs to* $\cap\mathcal{H}$*, then*

$$\mathcal{N}_{wmp}(\emptyset) = \mathcal{R}(\Phi_{\mathcal{H}}^{wBHL}) \text{ and } \bar{\Phi}_{\mathcal{H}}^{wBHL} = \bar{\Phi}(\mathcal{R}_{wmp}) \ .$$

In particular, the first step of the classical min-p procedure is equivalent to the practical procedure introduced by Baraud, Huet and Laurent [2] in the aggregated tests framework.
Note furthermore that because Bonferroni, Holm and min-p multiple testing procedures have a FWER controlled by the prescribed level α, their wFWER is also controlled by α, and so is the first kind error rate of the corresponding aggregated tests.

3.2. *From Separation Rates to Family Wise Separation Rates.* Let d be a distance on $\mathcal{P}$. For any P in $\mathcal{P}$ and any subset $\mathcal{Q}$ of $\mathcal{P}$, let

$$d\left(P, \mathcal{Q}\right) := \inf_{Q \in \mathcal{Q}} d(P, Q) \ .$$

Separation rates for aggregated tests. Separation rates are second kind error type quality criteria of a test of $H_0 \subset \mathcal{P}$ against $\mathcal{P} \setminus H_0$. Because $\mathcal{P}$ is in general too large to define separation rates over the whole set $\mathcal{P}$ properly, particularly in nonparametric frameworks, these quantities are first defined on a subset $\mathcal{Q}$ of $\mathcal{P}$. The question of adaptivity with respect to $\mathcal{Q}$ can then be treated. More precisely, we use the following definition due to Baraud [1], which can be viewed as a non-asymptotic version of Ingster's work [14].

DEFINITION 1. Given α and β in $(0,1)$, a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, and a test $\bar{\Phi}$ of the null hypothesis H_0, the uniform separation rate of $\bar{\Phi}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$\mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, H_0\right) = \inf\{r > 0 \quad / \quad \sup_{P \in \mathcal{Q}\ /\ d(P, H_0) \geq r} P(\bar{\Phi} = 0) \leq \beta\} \ .$$

Notice that when $H_0 \subset \cap\mathcal{H}$,

$$\mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, H_0\right) \geq \mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, \cap\mathcal{H}\right) \ .$$

The corresponding minimax separation rate over $\mathcal{Q}$ with prescribed level α and second kind error rate β is defined as

$$m\mathrm{SR}_d^{\alpha,\beta}\left(\mathcal{Q}, H_0\right) = \inf_{\bar{\Phi}\ /\ \mathrm{ER}\left(\bar{\Phi}, H_0\right) \leq \alpha} \mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, H_0\right) \ ,$$

where the infimum is taken over all possible level α tests.
A level α test $\bar{\Phi}$ is then said to be minimax over $\mathcal{Q}$ if $\mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, H_0\right)$ achieves $m\mathrm{SR}_d^{\alpha,\beta}\left(\mathcal{Q}, H_0\right)$, possibly up to a multiplicative constant depending on α and β.
Finally, it is said to be adaptive in the minimax sense over a collection of classes $\mathcal{Q}$ if $\mathrm{SR}_d^{\beta}\left(\bar{\Phi}, \mathcal{Q}, H_0\right)$ achieves, or nearly achieves, $m\mathrm{SR}_d^{\alpha,\beta}\left(\mathcal{Q}, H_0\right)$, for all the classes $\mathcal{Q}$ of the considered collection simultaneously, without knowing in advance to which class the distribution P belongs.

Weak Family Wise Separation Rates for multiple tests. Let us now consider a multiple testing procedure $\mathcal{R}$ and the corresponding aggregated test $\bar{\Phi}(\mathcal{R})$. Given β in $(0,1)$ and a class $\mathcal{Q} \subset \mathcal{P}$, according to Definition 1, the uniform separation rate of $\bar{\Phi}(\mathcal{R})$ over $\mathcal{Q}$ with prescribed second kind error rate β and distance d is defined by:

$$\mathrm{SR}_d^{\beta}\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right) = \inf\{r > 0 \ / \sup_{P \in \mathcal{Q} \ / \ d(P, \cap\mathcal{H}) \geq r} P\left(\mathcal{R} = \emptyset\right) \leq \beta\} \ .$$

This notion is closely related to the maximin optimality criterion considered by Romano, Shaikh and Wolf [22, Theorem 4.1], which consists in maximizing $\inf_{P \in \mathcal{Q} \subset \mathcal{P} \setminus \cap\mathcal{H}} P\left(\mathcal{R} \neq \emptyset\right)$.

Following the idea of the definition of the weak Family Wise Error Rate wFWER of $\mathcal{R}$, which is in fact equal to the first kind error rate of $\bar{\Phi}(\mathcal{R})$ for the null hypothesis $\cap\mathcal{H}$, a natural idea would be to define a notion of weak Family Wise Separation Rate as $\mathrm{SR}_d^{\beta}\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right)$. However, such a definition would not be satisfactory, as it could not be directly related to a stronger notion of Family Wise Separation Rate, and would not provide a complete minimax theory. One of the main cause of this is that "$d\left(P, \cap\mathcal{H}\right) \geq r$" (for $r > 0$) does not really define a suitable alternative to "$P \in \cap\mathcal{H}$" in the multiple testing philosophy, and should rather be replaced by "$\exists H \in \mathcal{H} \ / \ \ d\left(P, H\right) \geq r$".

This leads us to consider the set of false hypotheses under P at least at distance r from P, that is

$$\mathcal{F}_r(P) = \{ H \in \mathcal{H} \ / \ d(P, H) \geq r \} \ \ ,$$

which can be visualized on Figure 1.

From this set of false hypotheses under P at least at distance r from P, we can introduce the following definition.

DEFINITION 2. Given β in $(0,1)$ and a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, the weak Family Wise Separation Rate of a multiple test $\mathcal{R}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$w\mathrm{FWSR}_d^{\beta}\left(\mathcal{R}, \mathcal{Q}\right) = \inf\left\{ r > 0 \ / \sup_{P \in \mathcal{Q} \ / \ \mathcal{F}_r(P) \neq \emptyset} P\left(\mathcal{R} = \emptyset\right) \leq \beta \right\} \ .$$

This novel notion is however related to $\mathrm{SR}_d^{\beta}\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right)$, thanks to the following results.

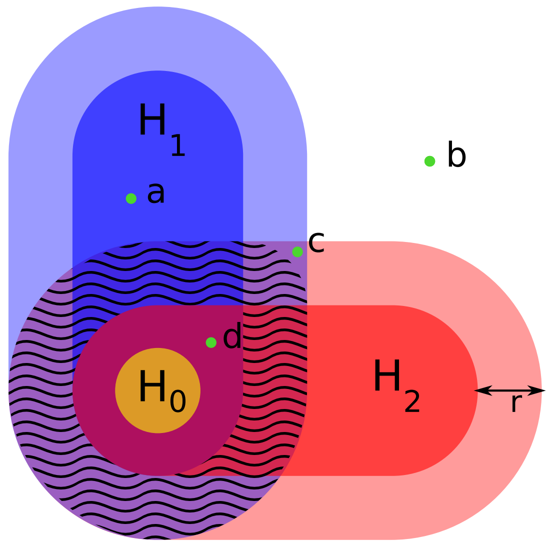

Figure 1. *Visualization of a multiple testing problem with two hypotheses H_1 and H_2, which are represented with darker colors. Their r-neighborhoods are of lighter shade. The r-neighborhood of $H_1 \cap H_2$ is hatched. A hypothesis H_0 is strictly included in $H_1 \cap H_2$. Point a corresponds to a distribution P such that $\mathcal{T}(P) = \{H_1\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_2\}$. Point b corresponds to a distribution P such that $\mathcal{T}(P) = \emptyset$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \{H_1, H_2\}$. Point c corresponds to a distribution P such that $\mathcal{T}(P) = \emptyset$, $\mathcal{F}(P) = \{H_1, H_2\}$, $\mathcal{F}_r(P) = \emptyset$ but $d(P, H_1 \cap H_2) \geq r$. Point d corresponds to a distribution P such that $\mathcal{T}(P) = \{H_1, H_2\}$ and $\mathcal{F}(P) = \mathcal{F}_r(P) = \emptyset$ but $P \notin H_0$.*

Proposition 3. *For any subset $\mathcal{Q}$ of $\mathcal{P}$ and β in $(0, 1)$,*

$$w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) \leq \mathrm{SR}_d^\beta\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right) \ .$$

Moreover, if the collection of hypotheses $\mathcal{H}$ and the distance d satisfy

$$\forall r > 0, \quad d(P, \cap\mathcal{H}) \geq r \quad \text{if and only if} \quad \mathcal{F}_r(P) \neq \emptyset \ , \tag{9}$$

then for every β in $(0, 1)$, and every class $\mathcal{Q} \subset \mathcal{P}$,

$$w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) = \mathrm{SR}_d^\beta\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right) \ .$$

Looking at Figure 1, the first inequality seems natural: for example, the point c is considered in $\mathrm{SR}_d^\beta\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right)$ but not in $w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$. It is therefore more difficult to control $\mathrm{SR}_d^\beta\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right)$ than $w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$. Note that if the collection of hypotheses $\mathcal{H}$ is closed (under intersection), that is

$$H \in \mathcal{H} \text{ and } H' \in \mathcal{H} \Rightarrow H \cap H' \in \mathcal{H} \ ,$$

condition (9) is always satisfied.
Furthermore, as another example, in the Gaussian regression framework considered in Section 2.3, if $\mathcal{H} = \{H_i, i = 1, ..., n\}$ where for every i in $\{1, \ldots, n\}$,

$$H_i = \{ P_{\mathbf{f}} \ / \ f_i = 0 \} \quad ,$$

and if d is taken as $d = d_\infty$, with

$$d_\infty(P_{\mathbf{f}}, P_{\mathbf{g}}) = \|\mathbf{f} - \mathbf{g}\|_\infty = \max_{i=1,\ldots,n} |f_i - g_i| \quad , \tag{10}$$

then condition (9) is also satisfied. But this is not true when using any other distance d_s for $s \geq 1$ defined by

$$d_s(P_{\mathbf{f}}, P_{\mathbf{g}}) = (\sum_{i=1}^{n} |f_i - g_i|^s)^{1/s} \quad . \tag{11}$$

See also Figure 1 drawn with d_2.

Family Wise Separation Rate for multiple tests. We now introduce a stronger notion of Family Wise Separation Rate, which defines a new second kind error type quality criterion for multiple testing procedures. It allows us to develop a minimax approach in the multiple testing framework, by bringing it closer to the well developed minimax theory for classical tests of a single null hypothesis.

Definition 3. Given β in $(0, 1)$ and a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, the Family Wise Separation Rate of a multiple test $\mathcal{R}$ over $\mathcal{Q}$ with prescribed second kind error rate β is defined by

$$\begin{aligned} \mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) &= \inf\left\{ r > 0 \ / \sup_{P \in \mathcal{Q}} P\left(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \emptyset\right) \leq \beta \right\} \\ &= \inf\left\{ r > 0 \ / \inf_{P \in \mathcal{Q}} P\left(\mathcal{F}_r(P) \subset \mathcal{R}\right) \geq 1 - \beta \right\} \quad . \end{aligned}$$

By definition, $\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q})$ is monotonous in $\mathcal{R}$, i.e. if $\mathcal{R} \subset \mathcal{R}'$ a.s., then

$$\mathrm{FWSR}_d^\beta\left(\mathcal{R}', \mathcal{Q}\right) \leq \mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) \quad . \tag{12}$$

The Family Wise Separation Rate is naturally a stronger quality criterion than the weak Family Wise Separation Rate, as stated in the following result.

Proposition 4. *For any subset $\mathcal{Q}$ of $\mathcal{P}$ and β in $(0, 1)$,*

$$w\mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) \leq \mathrm{FWSR}_d^\beta(\mathcal{R}, \mathcal{Q}) \quad .$$

Let us now introduce the corresponding minimax approach for multiple tests.

DEFINITION 4. Given α and β in $(0,1)$, a class of probability distributions $\mathcal{Q} \subset \mathcal{P}$, the minimax Family Wise Separation Rate over $\mathcal{Q}$ with prescribed FWER α and prescribed second kind error rate β is defined by

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) = \inf_{\mathcal{R}\ /\ \mathrm{FWER}(\mathcal{R}) \leq \alpha} \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q}) \quad ,$$

where the infimum is taken over all possible multiple tests with a FWER controlled by α.
A multiple test $\mathcal{R}$, whose FWER is controlled by α, is then said to be minimax over $\mathcal{Q}$ if $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ achieves $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q})$, possibly up to a multiplicative constant depending on α and β.
Finally, it is said to be adaptive in the minimax sense over a collection of classes $\mathcal{Q}$ if $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$ achieves, or nearly achieves, $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q})$, for all the classes $\mathcal{Q}$ of the considered collection simultaneously, without knowing in advance which class the distribution P belongs to.

This minimax approach can be linked, in certain cases, to the classical minimax theory for the tests of a single null hypothesis.

THEOREM 5. *If the distance d and the collection of hypotheses $\mathcal{H}$ satisfy* (9)*, for any subset $\mathcal{Q}$ of $\mathcal{P}$ and β in* $(0,1)$*,*

$$m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q}) \geq m\mathrm{SR}_d^{\alpha,\beta}(\mathcal{Q}, \cap\mathcal{H}) \quad . \tag{13}$$

The above result is not surprising: one can indeed think that testing multiple hypotheses is more difficult than testing a single hypothesis.
It directly gives lower bounds for the minimax Family Wise Error Rate over some classes $\mathcal{Q}$ by using the abundant literature on classical minimax testing. As proved by Theorem 8 in the following, note that some of these lower bounds are tight in some particular cases of Gaussian regression framework. But note also that condition (9) in Theorem 5 is necessary. Indeed, Theorem 7 exhibits counterexamples, not satisfying (9), for which the inequality (13) does not hold.

3.3. *Minimax Family Wise Separation Rates in the Gaussian regression framework.* In this section, we study in detail some (minimax) Family Wise Separation Rates in the classical Gaussian regression framework presented in Section 2.3.

3.3.1. *Example of failure of condition* (9) . We first focus on the collection

$$\mathcal{H} = \{H_{S_i},\ i = 1, ..., n\}\ ,$$

with $S_i = \mathrm{Vect}(e_i)$ and $H_{S_i} = \{P_{\mathbf{f}}\ /\ f_i = 0\} = \{P_{\mathbf{f}}\ /\ \mathbf{f} \in S_i^{\perp}\}$. Notice that when considering the distance $d = d_2$ as defined in (11) for instance, condition (9) fails.
In this framework, Baraud [1] studies the minimax Separation Rates for the null hypothesis $H_0 = \cap\mathcal{H} = \{P_0\}$ with $d = d_2$, over the classes of alternatives $\mathcal{Q} = \mathcal{P}_k$ defined, for any integer $k \leq n$, by

$$\mathcal{P}_k = \{\, P_{\mathbf{f}}\ /\ |\mathbf{f}|_0 \leq k \,\}\ , \tag{14}$$

where $|\mathbf{f}|_0$ is the number of non zero coefficients in $\mathbf{f}$. He proves in particular that for α and β in $(0, 1)$ such that $\alpha + \beta \leq 0.5$ and $k \geq 1$,

$$m\mathrm{SR}_{d_2}^{\alpha,\beta}\,(\mathcal{P}_k, H_0) \geq \sigma \left(k \ln\left(1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}}\right)\right)^{1/2}\ , \tag{15}$$

and that this lower bound is tight. Baraud, Huet and Laurent [2] then build aggregated tests that are adaptive, over a collection of classes $\mathcal{P}_k$ that is, able to achieve this rate without knowing the value of k for which P belongs to $\mathcal{P}_k$. In their case, σ^2 is not assumed to be known anymore. Moreover Laurent, Loubes, Marteau [18] further study the case of heteroscedasticity.
From this lower bound of Baraud, one can deduce the following theorem, whose proof is given in the last section.

THEOREM 6. *For any α, β in $(0, 1)$ such that $\alpha + \beta \leq 0.5$, for any s in $[1, \infty]$, for any k in $\{1, \dots, n\}$,*

$$m\mathrm{FWSR}_{d_s}^{\alpha,\beta}\,(\mathcal{P}_k) \geq \sigma\sqrt{\ln\left(1 + n\right)}\ .$$

Let us now prove that this lower bound is achieved. To do so, let us consider for any $i = 1 \dots n$, the p-value p_i corresponding to the test that rejects H_{S_i} when $T_i = |X_i|\sigma^{-1} > F^{-1}(1 - \alpha/2)$, where F is here the c.d.f. of a standard Gaussian distribution. Notice that since the Gaussian distribution is continuous, the three tests, as defined in Lemma 1, are identical, that is:

$$\mathbb{1}_{\{T_i > F^{-1}(1-\alpha/2)\}} = \mathbb{1}_{\left\{T_i > F_-^{-1}(1-\alpha/2)\right\}} = \mathbb{1}_{\{p_i \leq \alpha\}}\ .$$

THEOREM 7. *Let α in $(0, 1)$, and $\mathcal{R}$ be one of the four multiple testing procedures $\mathcal{R}_{Bonf}, \mathcal{R}_{Holm}, \mathcal{R}_{mp}$ and $\mathcal{R}(\Phi_{\mathcal{H}}^{BHL})$, based on the p-values p_i defined* *above, such that* $\mathrm{FWER}(\mathcal{R}) \le \alpha$*. Then for all* s *in* $[1, \infty]$*,* k *in* $\{1, \dots, n\}$*, and* β *in* $(0, 1)$*,*

$$\mathrm{FWSR}^{\beta}_{d_s}(\mathcal{R}, \mathcal{P}_k) \le \sigma\left(\sqrt{2\ln\left(\frac{k}{2\beta}\right)} + \sqrt{2\ln\left(\frac{2n}{\alpha}\right)}\right) \quad .$$

Comments.
(i) This proves that the four considered multiple testing procedures are minimax over the classes $\mathcal{P}_k$ with a Family Wise Separation Rate of order $\sigma(\ln(n))^{1/2}$, up to a multiplicative constant. Since the considered multiple tests do not depend on the value of k, they can be said adaptive in the minimax sense over all the classes $\mathcal{P}_k$, for $k = 1 \dots n$, simultaneously.
(ii) This also proves at the same time that the minimax Family Wise Separation Rate over $\mathcal{P}_k$ is of order $\sigma(\ln(n))^{1/2}$, which matches the minimax Separation Rate $m\mathrm{SR}^{\alpha,\beta}_{d_\infty}(\mathcal{P}_k, H_0)$ (see [17]).
More surprisingly, notice that when considering for instance the d_2 distance, the minimax Separation Rate $m\mathrm{SR}^{\alpha,\beta}_{d_2}(\mathcal{P}_k, H_0)$ is of order $\sigma n^{\gamma/2}\ln(n)^{1/2}$ when k is proportional to n^γ for $\gamma \in (0, 1/2)$ (see (15)), which is much larger than the above minimax Family Wise Separation Rate. This leads to think that, for this distance, performing a multiple testing procedure is easier than performing a test of a single hypothesis. This example thus shows the importance in Theorem 5 of condition (9), which is not satisfied when the distance d_2 is considered. This condition can therefore be viewed as a guarantee that performing a multiple testing procedure is more difficult than performing a test of a single hypothesis.

3.3.2. *Closed collection of hypotheses example.* For any i in $\{1, \dots, n\}$, let $\bar{S}_i = \mathrm{Vect}(e_1, \dots, e_i)$, so $H_{\bar{S}_i} = \{P_{\mathbf{f}} \ / \ f_1 = \dots = f_i = 0\} = \{P_{\mathbf{f}} \ / \ \Pi_{\bar{S}_i}(\mathbf{f}) = 0\}$. We now focus on the closed collection of hypotheses

$$\mathcal{H} = \{H_{\bar{S}_i},\ i = 1, ..., n\} \ ,$$

which satisfies $\cap\mathcal{H} = \{P_0\} = H_0$.
Let us here consider again the classes $\mathcal{P}_k$ defined by (14).
Since the above collection of hypotheses $\mathcal{H}$ is closed, condition (9) is always satisfied, and from Theorem 5, we deduce that for any distance d,

$$m\mathrm{FWSR}^{\alpha,\beta}_{d}(\mathcal{P}_k) \ge m\mathrm{SR}^{\alpha,\beta}_{d}(\mathcal{P}_k, H_0) \quad .$$

In particular, for $d = d_2$, from (15), the following lower bound is easily derived: for α and β in $(0, 1)$ such that $\alpha + \beta \le 0.5$, for k in $\{1, \dots, n\}$,

$$m\mathrm{FWSR}^{\alpha,\beta}_{d_2}(\mathcal{P}_k) \ge \sigma\left(k\ln\left(1 + \frac{n}{k^2} \vee \sqrt{\frac{n}{k^2}}\right)\right)^{1/2} \quad . \tag{16}$$

We now introduce a multiple testing procedure, which does not depend on the knowledge of k and whose Family Wise Separation Rate over $\mathcal{P}_k$ however achieves this lower bound, up to possible multiplicative constants.
As in the above section, let us consider again for any i in $\{1,\ldots,n\}$, the p-value p_i associated, thanks to Lemma 1, with the single test that rejects the null hypothesis $H_i = \{P_{\mathbf{f}} \ / \ f_i = 0\}$ when $T_i = |X_i|\sigma^{-1}$ takes large values. We then introduce the multiple test:

$$\bar{\mathcal{R}} = \{H_{\bar{S}_i} \ / \ \min_{j \leq i} p_j \leq \alpha/n\} \ . \tag{17}$$

As the c.d.f. F of the standard Gaussian distribution is continuous, notice that

$$\begin{aligned} \bar{\mathcal{R}} &= \left\{H_{\bar{S}_i} \ / \ \max_{j \leq i} T_j > F^{-1}\left(1 - \frac{\alpha}{2n}\right)\right\} \\ &= \left\{H_{\bar{S}_i} \ / \ \max_{j \leq i} T_j > F_{-}^{-1}\left(1 - \frac{\alpha}{2n}\right)\right\} \ . \end{aligned}$$

This procedure corresponds to a particular case of the variant of the closure method of [21] introduced by Romano and Wolf in [24, Algorithm 1 (idealized step-down method)] and [24, Theorem 1], when critical values satisfy a monotonicity assumption. In the notation of Romano and Wolf, here, $T_{n,i} = \max_{j \leq i} T_j$ and $d_{n,\{1,\ldots,i\}} = F^{-1}\left(1 - \frac{\alpha}{2n}\right)$ for all i in $\{1,\ldots,n\}$.

THEOREM 8. *Given α in $(0,1)$, let $\bar{\mathcal{R}}$ be the multiple test defined in (17). Then*

$$\mathrm{FWER}(\bar{\mathcal{R}}) \leq \alpha \ ,$$

and for any k in $\{1,\ldots,n\}$, β in $(0, 0.5)$,

$$\mathrm{FWSR}_{d_2}^{\beta}\left(\bar{\mathcal{R}}, \mathcal{P}_k\right) \leq \sigma\sqrt{k}\left(\sqrt{-2\ln(2\beta)} + \sqrt{2\ln(2n/\alpha)}\right) \ .$$

Comment. For k proportional to n^γ for $\gamma \in [0, 1/2)$, notice that this upper bound coincides with the lower bound obtained in (16). Hence we can conclude that, in this case at least, $m\mathrm{FWSR}_{d_2}^{\alpha,\beta}(\mathcal{P}_k)$ is of order $\sigma(n^\gamma \ln n)^{1/2}$, and that the multiple test $\bar{\mathcal{R}}$ defined by (17) is minimax adaptive over the considered classes. Notice moreover that there is here no price to pay for adaptation, and that such a phenomenon is rather rarely observed in minimax adaptive testing problems: up to our knowledge, only three cases are identified in [8], [9], and [18].

4. Proofs.

4.1. *Basic properties of the c.d.f.'s.* We give in the following lemma some basic properties of the càdlàg and càglàd c.d.f.'s, as defined in Section 2. As most of these properties are well-known and easy to prove, the proof of Lemma 9 is postponed to Appendix.

LEMMA 9. *For some random variable T, let F be its càdlàg cumulative distribution function (c.d.f.), F^{-1} the càglàd generalized inverse function of F, F_- its càglàd c.d.f., and F_-^{-1} the càdlàg generalized inverse function of F_-, that is:*

- $\forall t \in \mathbb{R}$, $F(t) = P(T \leq t)$,
- $\forall u \in [0,1]$, $F^{-1}(u) = \inf\{t,\ F(t) \geq u\}$,
- $\forall t \in \mathbb{R}$, $F_-(t) = P(T < t)$,
- $\forall u \in [0,1]$, $F_-^{-1}(u) = \sup\{t,\ F_-(t) \leq u\}$.

Then, for any t in $\mathbb{R}$ and any fixed u in $[0,1]$,

1. $F^{-1}(F(t)) \leq t \leq F_-^{-1}(F_-(t))$,
2. $F_-\left(F_-^{-1}(u)\right) \leq u \leq F\left(F^{-1}(u)\right)$,
3. $F^{-1}(u) \leq t \Leftrightarrow u \leq F(t)$, *and* $F_-^{-1}(u) \geq t \Leftrightarrow u \geq F_-(t)$,
4. $F_-(F^{-1}(u)) \leq u \leq F(F_-^{-1}(u))$,
5. $F^{-1}(u) \leq F_-^{-1}(u)$,
6. $F^{-1}(u) < F_-^{-1}(u) \Rightarrow F_-\left(F_-^{-1}(u)\right) = F\left(F^{-1}(u)\right) = u$,
7. *almost surely in T*, $\mathbb{1}_{\{F_-(T)=u\}} = \mathbb{1}_{\left\{T=F_-^{-1}(u)\right\}}\mathbb{1}_{\left\{F_-\left(F_-^{-1}(u)\right)=u\right\}}$.

COROLLARY 10. *With the notations of Lemma 1,*

$$\begin{cases} \phi_p = \mathbb{1}_{\left\{T>F_-^{-1}(1-\alpha)\right\}} = \mathbb{1}_{\{T>F^{-1}(1-\alpha)\}} \text{ a.s. if } F_-(F_-^{-1}(1-\alpha)) < 1-\alpha\ , \\ \phi_p = \mathbb{1}_{\left\{T\geq F_-^{-1}(1-\alpha)\right\}} \text{ a.s. if } F_-(F_-^{-1}(1-\alpha)) = 1-\alpha\ . \end{cases}$$

PROOF. We easily deduce from point 3 and point 7 of Lemma 9 above that:

$$\begin{aligned} \phi_p &= \mathbb{1}_{\{p(T)\leq\alpha\}} \\ &= \mathbb{1}_{\{F_-(T)\geq 1-\alpha\}} \\ &= \mathbb{1}_{\{F_-(T)>1-\alpha\}} + \mathbb{1}_{\{F_-(T)=1-\alpha\}} \\ &= \mathbb{1}_{\left\{T>F_-^{-1}(1-\alpha)\right\}} + \mathbb{1}_{\{F_-(T)=1-\alpha\}} \\ &= \mathbb{1}_{\left\{T>F_-^{-1}(1-\alpha)\right\}} + \mathbb{1}_{\left\{T=F_-^{-1}(1-\alpha)\right\}}\mathbb{1}_{\left\{F_-\left(F_-^{-1}(1-\alpha)\right)=1-\alpha\right\}} \text{ a.s.} \end{aligned}$$

The result is thus finally derived from point 6 of Lemma 9 which gives that when $F_-\left(F_-^{-1}(1-\alpha)\right) < 1-\alpha$, $F_-^{-1}(1-\alpha) = F^{-1}(1-\alpha)$. □

4.2. *Proof of Lemma 1.* Let us assume that P is a probability distribution in H_0. From point 2 and point 4 of Lemma 9, we easily obtain that

$$P(\phi = 1) = 1 - F(F^{-1}(1-\alpha)) \leq \alpha, \text{ and } P(\phi_- = 1) = 1 - F(F_-^{-1}(1-\alpha)) \leq \alpha \ .$$

Therefore, ϕ and ϕ_- are both of level α.
Moreover, from Corollary 10, if $F_-\left(F_-^{-1}(1-\alpha)\right) < 1-\alpha$, then $\phi_p = \phi$ a.s., so ϕ_p is also of level α. If $F_-\left(F_-^{-1}(1-\alpha)\right) = 1-\alpha$, then $\phi_p = \mathbb{1}_{\left\{T \geq F_-^{-1}(1-\alpha)\right\}}$ a.s., so

$$P(\phi_p = 1) = 1 - F_-(F_-^{-1}(1-\alpha)) = \alpha \ ,$$

and ϕ_p is still of level α.
Now, notice that the three considered tests ϕ, ϕ_- and ϕ_p are monotonous in the following sense: for any $\alpha < \alpha'$,

$$\begin{cases} T > F^{-1}(1-\alpha)\} \Rightarrow T > F^{-1}(1-\alpha') \ , \\ T > F_-^{-1}(1-\alpha)\} \Rightarrow T > F_-^{-1}(1-\alpha') \ , \\ p(T) \leq \alpha \Rightarrow p(T) \leq \alpha' \ . \end{cases}$$

Hence, the p-values associated with ϕ, ϕ_- and ϕ_p are respectively defined by $\inf\{\alpha \ / \ T > F^{-1}(1-\alpha)\}$, $\inf\{\alpha \ / \ T > F_-^{-1}(1-\alpha)\}$, and $\inf\{\alpha \ / \ p(T) \leq \alpha\}$. From point 3 of Lemma 9, it is obvious that $\inf\{\alpha \ / \ T > F_-^{-1}(1-\alpha)\} = \inf\{\alpha \ / \ 1 - F_-(T) < \alpha\} = \inf\{\alpha \ / \ p(T) < \alpha\}$, so the p-value associated with ϕ_- is equal to $p(T)$. In the same way, it is also obvious that the p-value associated with ϕ_p defined as $\inf\{\alpha \ / \ p(T) \leq \alpha\}$ is equal to $p(T)$.
As for the p-value associated with ϕ, it is not so clear. Let $\tilde{\alpha}$ denote this p-value.
First, if T is not an atom of the underlying distribution, F is continuous in T and $F_-(T) = F(T)$. If $\alpha > \tilde{\alpha}$, then $T > F^{-1}(1-\alpha)$. Since F is non decreasing, this gives, with point 2 of Lemma 9: $F(T) = F_-(T) \geq F(F^{-1}(1-\alpha)) \geq 1-\alpha$, therefore $\alpha \geq p(T)$. If $\alpha < \tilde{\alpha}$, then $T \leq F^{-1}(1-\alpha)$. Since F_- is also non decreasing, this gives, with point 4 of Lemma 9: $F_-(T) \leq F_-(F^{-1}(1-\alpha)) \leq 1-\alpha$, therefore $\alpha \leq p(T)$. Hence, if T is not an atom of the underlying distribution, $\tilde{\alpha} = p(T)$.
Next, if T is an atom, then F^{-1} is constant on the interval $(F_-(T), F(T)]$ and its value on this interval is T. If $\alpha > \tilde{\alpha}$ then $T > F^{-1}(1-\alpha)$, so $1-\alpha \leq F_-(T)$ i.e. $\alpha \geq p(T)$. If $\alpha < \tilde{\alpha}$, then, from point 5 of Lemma 9, $T \leq F^{-1}(1-\alpha) \leq F_-^{-1}(1-\alpha)$. From point 2 or 3 of Lemma 9, we deduce that $F_-(T) \leq 1-\alpha$, that is $\alpha \leq p(T)$. Therefore $\tilde{\alpha}$ is also equal to $p(T)$ in this case.
To conclude, (1) and (2) are finally easily deduced from point 5 and point 3 of Lemma 9.

4.3. *Proof of Proposition 2.* The first part of Proposition 2 is a straightforward consequence of (1). As for the min-p procedure, one can write:

$$\begin{aligned} P\left(\exists H \in \mathcal{H} \ / \ T_H > F_{H,-}^{-1}(1 - w_H u)\right) &= P\left(\exists H \in \mathcal{H} \ / \ w_H^{-1} p_H < u\right) \\ &= P\left(\min_{H \in \mathcal{H}} w_H^{-1} p_H < u\right) \\ &= F_-(u) \ , \end{aligned}$$

where F is the c.d.f. of $\min_{H\in\mathcal{H}} w_H^{-1} p_H$, which does not depend on P in $\cap\mathcal{H}$. Therefore, by definition, $u_\alpha = F_-^{-1}(\alpha)$. From (1) again, we finally derive the result.

4.4. *Proof of Proposition 3.* If $\mathcal{F}_r(P) \neq \emptyset$, then there exists H in $\mathcal{H}$ such that $d(P,H) \geq r$, i.e. such that for any Q in H, $d(P,Q) \geq r$. In particular this is true for every Q in $\cap\mathcal{H} \subset H$, so $d(P, \cap\mathcal{H}) \geq r$. Therefore

$$\sup_{P\in\mathcal{Q} \ / \ \mathcal{F}_r(P)\neq\emptyset} P(\mathcal{R} = \emptyset) \leq \sup_{P\in\mathcal{Q} \ / \ d(P,\cap\mathcal{H})\geq r} P(\mathcal{R} = \emptyset) \ .$$

Hence,

$$\begin{aligned} &\left\{ r > 0, \sup_{P\in\mathcal{Q} \ / \ d(P,\cap\mathcal{H})\geq r} P(\mathcal{R} = \emptyset) \leq \beta \right\} \\ &\qquad\qquad \subset \left\{ r > 0, \sup_{P\in\mathcal{Q} \ / \ \mathcal{F}_r(P)\neq\emptyset} P(\mathcal{R} = \emptyset) \leq \beta \right\} \ , \end{aligned}$$

which leads to the first inequality. But of course under condition (9), both sets are equal and the inequality becomes an equality.

4.5. *Proof of Proposition 4.* Noticing that for $r > 0$,

$$\begin{aligned} \sup_{P\in\mathcal{Q} \ / \ \mathcal{F}_r(P)\neq\emptyset} P\left(\mathcal{R} = \emptyset\right) &\leq \sup_{P\in\mathcal{Q} \ / \ \mathcal{F}_r(P)\neq\emptyset} P\left(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \emptyset\right) \\ &\leq \sup_{P\in\mathcal{Q}} P\left(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \emptyset\right) \ , \end{aligned}$$

we derive that

$$\begin{aligned} &\left\{ r > 0, \sup_{P\in\mathcal{Q}} P\left(\mathcal{F}_r(P) \cap (\mathcal{H} \setminus \mathcal{R}) \neq \emptyset\right) \leq \beta \right\} \\ &\qquad\qquad \subset \left\{ r > 0, \sup_{P\in\mathcal{Q} \ / \ \mathcal{F}_r(P)\neq\emptyset} P\left(\mathcal{R} = \emptyset\right) \leq \beta \right\} \ , \end{aligned}$$

which gives the result.

4.6. *Proof of Theorem 13.* Since for any multiple testing procedure $\mathcal{R}$, $\mathrm{FWER}(\mathcal{R}) \geq w\mathrm{FWER}(\mathcal{R}) = \mathrm{ER}(\bar{\Phi}(\mathcal{R}), \cap\mathcal{H})$ by (8), one has that

$$m\mathrm{FWSR}_d^{\alpha,\beta}\left(\mathcal{Q}\right) \geq \inf_{\mathcal{R}\ /\ \mathrm{ER}\left(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H}\right)\leq\alpha} \mathrm{FWSR}_d^{\beta}\left(\mathcal{R}, \mathcal{Q}\right) \ .$$

By Proposition 4, one as that

$$m\mathrm{FWSR}_d^{\alpha,\beta}\left(\mathcal{Q}\right) \geq \inf_{\mathcal{R}\ /\ \mathrm{ER}\left(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H}\right)\leq\alpha} w\mathrm{FWSR}_d^{\beta}\left(\mathcal{R}, \mathcal{Q}\right) \ .$$

By condition (9), this is equivalent to

$$m\mathrm{FWSR}_d^{\alpha,\beta}\left(\mathcal{Q}\right) \geq \inf_{\mathcal{R}\ /\ \mathrm{ER}\left(\bar{\Phi}(\mathcal{R}),\cap\mathcal{H}\right)\leq\alpha} \mathrm{SR}_d^{\beta}\left(\bar{\Phi}(\mathcal{R}), \mathcal{Q}, \cap\mathcal{H}\right) \ ,$$

which allows to conclude as a $\bar{\Phi}(\mathcal{R})$ is a particular single test of $\cap\mathcal{H}$.

4.7. *Proof of Theorem 6.* First notice that for any $\mathcal{Q} \subset \mathcal{A}$,

$$\left\{ r > 0,\ \sup_{P\in\mathcal{A}} P(\mathcal{F}_r(P) \cap (\mathcal{H}\setminus\mathcal{R}) \neq \emptyset) \leq \beta \right\} \\ \subset \left\{ r > 0,\ \sup_{P\in\mathcal{Q}} P(\mathcal{F}_r(P) \cap (\mathcal{H}\setminus\mathcal{R}) \neq \emptyset) \leq \beta \right\} \ ,$$

and therefore $\mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{A}) \geq \mathrm{FWSR}_d^{\beta}(\mathcal{R}, \mathcal{Q})$.
This also implies that $m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{A}) \geq m\mathrm{FWSR}_d^{\alpha,\beta}(\mathcal{Q})$.
Since $\mathcal{P}_1 \subset \mathcal{P}_k$ for every k in $\{1, \ldots, n\}$, then

$$m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) \geq m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_1) = m\mathrm{FWSR}_{d_\infty}^{\alpha,\beta}(\mathcal{P}_1) \ .$$

The last equality comes from the fact that for every distance d_s defined by (11) and (10) respectively, and for any $P_{\mathbf{f}}$ in $\mathcal{P}_1$ and any $i = 1, ..., n$,

$$d_s(P_{\mathbf{f}}, H_{S_i}) = d_\infty(P_{\mathbf{f}}, H_{S_i}) = d_2(P_{\mathbf{f}}, H_{S_i}) \ .$$

By Theorem 13, since condition (9) holds for d_∞,

$$m\mathrm{FWSR}_{d_s}^{\alpha,\beta}(\mathcal{P}_k) \geq m\mathrm{SR}_{d_\infty}^{\alpha,\beta}\left(\mathcal{P}_1, \cap\mathcal{H}\right) = m\mathrm{SR}_{d_2}^{\alpha,\beta}\left(\mathcal{P}_1, \cap\mathcal{H}\right) \ .$$

We finally conclude thanks to the lower bound (15).

4.8. *Proof of Theorem 7.* By construction, $\mathcal{R}_{Bonf} \subset \mathcal{R}_{Holm} \subset \mathcal{R}_{mp}$. By Proposition 2, we also have that $\mathcal{R}_{Bonf} = \mathcal{N}_{Holm}(\emptyset) \subset \mathcal{N}_{mp}(\emptyset) = \mathcal{R}(\Phi_{\mathcal{H}}^{BHL})$. It is therefore sufficient to upper bound $\mathrm{FWSR}_{d_s}^{\beta}(\mathcal{R}_{Bonf}, \mathcal{P}_k)$ by (12).
Therefore, the aim here is to find a r_0 such that for any $r \geq r_0$ and for any $P_{\mathbf{f}}$ in $\mathcal{P}_k$,

$$P_{\mathbf{f}}\left(\mathcal{F}_r(P_{\mathbf{f}}) \subset \mathcal{R}_{Bonf}\right) \geq 1-\beta \ .$$

By independence, since $d_s(P, H_{S_i}) = \inf_{P_{\mathbf{g}} \in H_{S_i}} d_s(P_{\mathbf{f}}, P_{\mathbf{g}}) = |f_i|$,

$$\begin{aligned} P_{\mathbf{f}}\left(\mathcal{F}_r(P_{\mathbf{f}}) \subset \mathcal{R}_{Bonf}\right) &= P_{\mathbf{f}}\left(\forall i \text{ s.t. } d_s(P_{\mathbf{f}}, H_{S_i}) \geq r,\ H_{S_i} \in \mathcal{R}_{Bonf}\right) \\ &= P_{\mathbf{f}}\left(\forall i \text{ s.t. } |f_i| \geq r,\ p_i \leq \alpha/n\right) \\ &= \prod_{i \text{ s.t. } |f_i| \geq r} P_{\mathbf{f}}\left(p_i \leq \alpha/n\right) \ . \end{aligned}$$

Moreover, denoting by F the c.d.f. of ε, a standard Gaussian variable, recall that

$$p_i = 2F(-\sigma^{-1}|X_i|) = 2F\left(-\left|\frac{f_i}{\sigma} + \varepsilon_i\right|\right).$$

One can show easily that for all real numbers a, b

$$P\left(|a+\varepsilon| > b\right) \geq F(|a| - b) \ . \tag{18}$$

Hence

$$\begin{aligned} P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \mathcal{R}_{Bonf}) &= \prod_{i \text{ s.t. } |f_i| \geq r} P_{\mathbf{f}}\left(\left|\frac{f_i}{\sigma} + \varepsilon_i\right| \geq -F^{-1}\left(\frac{\alpha}{2n}\right)\right) \\ &\geq \prod_{i \text{ s.t. } |f_i| \geq r} \left(F\left(\frac{|f_i|}{\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right)\right) \\ &\geq \left(F\left(\frac{r}{\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right)\right)^{\#\mathcal{F}_r(P_{\mathbf{f}})} \ . \end{aligned}$$

Hence, $P_{\mathbf{f}}\left(\mathcal{F}_r(P_{\mathbf{f}}) \subset \mathcal{R}_{Bonf}\right) \geq 1-\beta$ if

$$r \geq \sigma\left(F^{-1}\left((1-\beta)^{1/\#\mathcal{F}_r(P_{\mathbf{f}})}\right) - F^{-1}\left(\frac{\alpha}{2n}\right)\right) \ ,$$

or if

$$r \geq \sigma\left(F^{-1}\left((1-\beta)^{1/\#\mathcal{F}_r(P_{\mathbf{f}})}\right) + F^{-1}\left(1-\frac{\alpha}{2n}\right)\right) \ .$$

Let us now recall the following bound on the tail of the Gaussian distribution:

$$\forall u > 0, \qquad 1 - F(u) \leq \frac{1}{2} e^{-u^2/2} \ .$$

This implies that

$$\forall u \in (0,1), \qquad F^{-1}(u) \leq \sqrt{-2\ln\left(2(1-u)\right)}\ . \tag{19}$$

Therefore, $P_{\mathbf{f}}\left(\mathcal{F}_r(P_{\mathbf{f}}) \subset \mathcal{R}_{Bonf}\right) \geq 1-\beta$ if

$$r \geq \sigma\left(\sqrt{-2\ln\left(2(1-(1-\beta)^{1/\#\mathcal{F}_r(P_{\mathbf{f}})})\right)} + \sqrt{-2\ln\left(\frac{\alpha}{2n}\right)}\right)\ .$$

Finally, we use that for any u in $(0,1)$ and x in $[0,1]$,

$$x(1-u) \leq 1-u^x\ .$$

This result is easily derived by introducing $g(u) = 1-u^x - x(1-u)$, and noticing that g' is negative and $g(1) = 0$.
We thus obtain that $P_{\mathbf{f}}\left(\mathcal{F}_r(P) \subset \mathcal{R}_{Bonf}\right) \geq 1-\beta$ if

$$r \geq \sigma\left(\sqrt{2\ln\left(\frac{\#\mathcal{F}_r(P_{\mathbf{f}})}{2\beta}\right)} + \sqrt{2\ln\left(\frac{2n}{\alpha}\right)}\right)\ .$$

This concludes the proof of the theorem, since $\#\mathcal{F}_r(P_{\mathbf{f}}) \leq k$ when P belongs to $\mathcal{P}_k$.

4.9. *Proof of Theorem 8.* Let us first prove that $\mathrm{FWER}(\bar{\mathcal{R}}) \leq \alpha$. For any $\mathbf{f}$, let i_0 be the largest integer i such that $f_1 = ... = f_i = 0$. Then,

$$\begin{aligned} P_{\mathbf{f}}(\bar{\mathcal{R}} \cap \mathcal{T}(P_{\mathbf{f}}) \neq \emptyset) &= P_{\mathbf{f}}(\exists i \leq i_0, \exists j \leq i,\ p_j \leq \alpha/n) \\ &\leq P_{\mathbf{f}}(\exists j \leq i_0,\ p_j \leq \alpha/n) \\ &\leq \alpha\ . \end{aligned}$$

Considering $d = d_2$, the goal is now to find a r_0 such that for any $r \geq r_0$ and for any $P_{\mathbf{f}}$ in $\mathcal{P}_k$,

$$P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) \geq 1-\beta\ .$$

Assume that $P_{\mathbf{f}}$ belongs to $\mathcal{P}_k$. Given $r > 0$,

$$\mathcal{F}_r(P_{\mathbf{f}}) = \left\{ H_{\bar{S}_i}\ /\ d_2(P_{\mathbf{f}}, H_{\bar{S}_i}) \geq r \right\} = \left\{ H_{\bar{S}_i}\ /\ \sum_{j\in\{1,\ldots,i\}} f_j^2 \geq r^2 \right\}\ .$$

Then, if $\sum_{j=1}^n f_j^2 < r^2$, $P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) = 1$. Otherwise, let i_0 be now the smallest integer in $\{1,\ldots,n\}$ such that $\sum_{j=1}^{i_0} f_j^2 \geq r^2$. As this sum has at most $i_0 \wedge k$ non zero terms, there exists j_0 in $\{1, \ldots, i_0\}$ such that $f_{j_0}^2 \geq r^2/(i_0 \wedge k) \geq r^2/k$. Furthermore,

$$P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) = P_{\mathbf{f}}\left(\forall i \text{ s.t. } \sum_{j \in \{1,\ldots,i\}} f_j^2 \geq r^2, \min_{j \in \{1,\ldots,i\}} p_j \leq \frac{\alpha}{n}\right) .$$

If $p_{j_0} \leq \alpha/n$, then $\min_{j=1,\ldots,i_0} p_j \leq \alpha/n$, and for every i in $\{1, \ldots, n\}$ such that $\sum_{j=1}^{i} f_j^2 \geq r^2$, one has that $i \geq i_0 \geq j_0$. Hence $\min_{j=1,\ldots,i} p_j \leq \alpha/n$. The event $\left\{\forall i \text{ s.t. } \sum_{j=1}^{i} f_j^2 \geq r^2, \min_{j=1,\ldots,i} p_j \leq \alpha/n\right\}$ thus contains the event $\{p_{j_0} \leq \alpha/n\}$. As a consequence, with the notations of the above proof of Theorem 7,

$$\begin{aligned} P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) &\geq P_{\mathbf{f}}\left(p_{j_0} \leq \alpha/n\right) \\ &\geq P_{\mathbf{f}}\left(2F(-\sigma^{-1}|X_{j_0}|) \leq \alpha/n\right) \\ &\geq P_{\mathbf{f}}\left(2F\left(-\left|\frac{f_{j_0}}{\sigma} + \varepsilon_i\right|\right) \leq \alpha/n\right) . \end{aligned}$$

By (18), it follows that

$$\begin{aligned} P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) &\geq F\left(\frac{|f_{j_0}|}{\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right) \\ &\geq F\left(\frac{r}{\sqrt{k}\sigma} + F^{-1}\left(\frac{\alpha}{2n}\right)\right) . \end{aligned}$$

Therefore, $P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) \geq 1 - \beta$ as soon as

$$F\left(\frac{r}{\sigma\sqrt{k}} + F^{-1}\left(\frac{\alpha}{2n}\right)\right) \geq 1 - \beta .$$

Finally, by (19), we derive that $P_{\mathbf{f}}(\mathcal{F}_r(P_{\mathbf{f}}) \subset \bar{\mathcal{R}}) \geq 1 - \beta$ as soon as

$$\frac{r}{\sigma\sqrt{k}} \geq \sqrt{-2\ln(2\beta)} + \sqrt{2\ln(2n/\alpha)} .$$

This concludes the proof.

Acknowledgments. This research was partly supported by the french Agence Nationale de la Recherche (ANR 2011 BS01 010 01 projet Calibration).

References.

[1] Yannick Baraud. Non-asymptotic minimax rates of testing in signal detection. *Bernoulli*, 8(5):577–606, 2002.

[2] Yannick Baraud, Sylvie Huet, and Béatrice Laurent. Adaptive tests of linear hypotheses by model selection. *The Annals of Statistics*, 31(1):225–251, 2003.

[3] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B*, 57(1):289–300, 1995.

[4] Olivier Collier and Arnak Dalalyan. Curve registration by nonparametric goodness-of-fit testing. *arXiv preprint arXiv:1104.4210*, 2011.

[5] Sandrine Dudoit and Mark J. Van Der Laan. *Multiple testing procedures with applications to genomics*. Springer, 2007.

[6] Magalie Fromont and Béatrice Laurent. Adaptive goodness-of-fit tests in a density model. *The Annals of Statistics*, 34(2):680–720, 2006.

[7] Magalie Fromont, Béatrice Laurent, Matthieu Lerasle, and Patricia Reynaud-Bouret. Kernels based tests with non-asymptotic bootstrap approaches for two-sample problems. *Journal of Machine Learning Research: Workshop and Conference Proceedings, COLT 2012*, 23:23.1–23.22, 2012.

[8] Magalie Fromont, Béatrice Laurent, and Patricia Reynaud-Bouret. Adaptive tests of homogeneity for a poisson process. *Annales de l'Institut Henri Poincaré Probab. Stat.*, 47(1):176–213, 2011.

[9] Magalie Fromont, Béatrice Laurent, and Patricia Reynaud-Bouret. The two-sample problem for poisson processes: Adaptive tests with a nonasymptotic wild bootstrap approach. *The Annals of Statistics*, 41(3):1431–1461, 2013.

[10] Christopher Genovese and Larry Wasserman. A stochastic process approach to false discovery control. *Annals of Statistics*, pages 1035–1061, 2004.

[11] Jelle J. Goeman and Aldo Solari. The sequential rejection principle of familywise error control. *The Annals of Statistics*, 38(6):3782–3810, 2010.

[12] Sture Holm. A simple sequentially rejective multiple test procedure. *Scandinavian journal of statistics*, 6(2):65–70, 1979.

[13] G. Hommel and T. Hoffmann. Controlled uncertainty. In *Multiple Hypothesenprüfung/Multiple Hypotheses Testing*, pages 154–161. Springer, 1988.

[14] Yuri I. Ingster. Asymptotically minimax testing for nonparametric alternatives I-II-III. *Mathematical Methods of Statististics*, 2:85–114, 171–189, 249–268, 1993.

[15] Yuri I. Ingster. Adaptive chi-square tests. *Journal of Mathematical Sciences*, 99(2):1110–1119, 2000.

[16] Edward L. Korn, James F. Troendle, Lisa M. McShane, and Richard Simon. Controlling the number of false discoveries: application to high-dimensional genomic data. *Journal of Statistical Planning and Inference*, 124(2):379–398, 2004.

[17] Béatrice Laurent, Jean-Michel Loubes, and Clément Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. `http://arxiv.org/abs/0912.2423v1`.

[18] Béatrice Laurent, Jean-Michel Loubes, and Clément Marteau. Non asymptotic minimax rates of testing in signal detection with heterogeneous variances. *Electronic Journal of Statistics*, 6:91–122, 2012.

[19] E. L. Lehmann and Joseph P. Romano. Generalizations of the familywise error rate. *The Annals of Statistics*, 33(3):1138–1154, 2005.

[20] E. L. Lehmann, Joseph P. Romano, and Juliet Popper Shaffer. On optimality of stepdown and stepup multiple test procedures. *The Annals of Statistics*, 33(3):1084–

1108, 2005.
[21] Ruth Marcus, Peritz Eric, and K. Ruben Gabriel. On closed testing procedures with special reference to ordered analysis of variance. *Biometrika*, 63(3):655–660, 1976.
[22] Joseph P. Romano, Azeem Shaikh, and Michael Wolf. Consonance and the closure method in multiple testing. *The International Journal of Biostatistics*, 7(1):1–25, 2011.
[23] Joseph P. Romano and Azeem M. Shaikh. Stepup procedures for control of generalizations of the familywise error rate. *The Annals of Statistics*, 34(4):1850–1873, 2006.
[24] Joseph P. Romano and Michael Wolf. Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association*, 100(469):94–108, 2005.
[25] Joseph P. Romano and Michael Wolf. Control of generalized error rates in multiple testing. *The Annals of Statistics*, 35(4):1378–1408, 2007.
[26] Joseph P. Romano and Michael Wolf. Balanced control of generalized error rates. *The Annals of Statistics*, 38(1):598–633, 2010.
[27] R. J. Simes. An improved bonferroni procedure for multiple tests of significance. *Biometrika*, 73(3):751–754, 1986.
[28] Emil Spjotvoll. On the optimality of some multiple comparison procedures. *The Annals of Mathematical Statistics*, 43(2):398–411, 1972.
[29] Vladimir G. Spokoiny. Adaptive hypothesis testing using wavelets. *The Annals of Statistics*, 24(6):2477–2498, 1996.

APPENDIX A: PROOF OF LEMMA 9

Let for every u in $[0,1]$, $\mathcal{F}_u = \{t,\ F(t) \geq u\}$ and $\mathcal{F}^-_u = \{t,\ F_-(t) \leq u\}$. Then $F^{-1}(u) = \inf \mathcal{F}_u$ and $F_-^{-1}(u) = \sup \mathcal{F}^-_u$. Note that if $\mathcal{F}_u = \emptyset$ then $u = 1$ and $F^{-1}(u) = +\infty$. Similarly if $\mathcal{F}^-_u = \emptyset$ then $u = 0$ and $F_-^{-1}(u) = -\infty$. By taking the limit, we also have that $F(+\infty) = F_-(+\infty) = 1$ and $F(-\infty) = F_-(-\infty) = 0$.
Let us now fix t in $\mathbb{R}$ and u in $[0,1]$.

1. It is clear that $t \in \mathcal{F}_{F(t)}$ and $t \in \mathcal{F}^-_{F_-(t)}$, so

$$F^{-1}\left(F(t)\right) = \inf \mathcal{F}_{F(t)} \leq t \leq \sup \mathcal{F}^-_{F_-(t)} = F_-^{-1}\left(F_-(t)\right) \quad .$$

2. By definition, if $\mathcal{F}_u \neq \emptyset$, there exists a decreasing sequence $(t_n)_{n\in\mathbb{N}}$ of elements in $\mathcal{F}_u$ such that $t_n \to_{n\to+\infty} F^{-1}(u)$. Since $F(t_n) \geq u$ for every n in $\mathbb{N}$ and F is càdlàg, then making n tend to $+\infty$, $F(F^{-1}(u)) \geq u$. If $\mathcal{F}_u = \emptyset$, $u = 1$, $F^{-1}(u) = +\infty$ and $F(F^{-1}(u)) = F(+\infty) = 1 = u$.
In the same way, if $\mathcal{F}^-_u \neq \emptyset$, there exists an increasing sequence $(t^-_n)_{n\in\mathbb{N}}$ of elements in $\mathcal{F}^-_u$ such that $t^-_n \to_{n\to+\infty} F_-^{-1}(u)$. Since $F_-(t^-_n) \leq u$ for every n in $\mathbb{N}$ and F_- is càglàd, then $F_-(F_-^{-1}(u)) \leq u$. If $\mathcal{F}^-_u = \emptyset$, $u = 0$, $F^{-1}(u) = -\infty$ and $F_-(F_-^{-1}(u)) = F_-(-\infty) = 0 = u$.

3. Assume that $u \le F(t)$. Then $t \in \mathcal{F}_u$ (and $\mathcal{F}_u \neq \emptyset$), so obviously $F^{-1}(u) = \inf \mathcal{F}_u \le t$.
Conversely, assume that $F^{-1}(u) \le t$. Since F is increasing, $F(F^{-1}(u)) \le F(t)$. From point 2 above, this implies that $F(t) \ge u$.
Assume now that $u \ge F_-(t)$. Then $t \in \mathcal{F}_u^-$ (and $\mathcal{F}_u^- \neq \emptyset$), so $F_-^{-1}(u) \ge t$.
Conversely, assume that $F_-^{-1}(u) \ge t$. Since F_- is also increasing, $F_-(F_-^{-1}(u)) \ge F_-(t)$, which leads with point 2 to $u \ge F_-(t)$.

4. First assume that $F^{-1}(u) \neq +/-\infty$. Then for every $n \ge 1$, $F^{-1}(u) - 1/n < \inf \mathcal{F}_u$, so $F^{-1}(u) - 1/n \notin \mathcal{F}_u$ and $F_-\left(F^{-1}(u) - 1/n\right) \le F\left(F^{-1}(u) - 1/n\right) < u$. Furthermore, F_- is càglàd, so taking the limit as n tends to $+\infty$ gives $F_-\left(F^{-1}(u)\right) \le u$. If $F^{-1}(u) = +\infty$, $u = 1$ and $F_-(F^{-1}(u)) = F_-(+\infty) = 1 = u$. If $F^{-1}(u) = -\infty$, $u = 0$ and $F_-(F^{-1}(u)) = F_-(-\infty) = 0 = u$.
In the same way, assume that $F_-^{-1}(u) \neq +/-\infty$. Then for every $n \ge 1$, $F_-^{-1}(u) + 1/n > \sup \mathcal{F}_u^-$, so $F_-^{-1}(u) + 1/n \notin \mathcal{F}_u^-$ and $F\left(F_-^{-1}(u) + 1/n\right) \ge F_-\left(F_-^{-1}(u) + 1/n\right) > u$. As F is càdlàg, taking the limit as n tends to $+\infty$ leads to $F\left(F_-^{-1}(u)\right) \ge u$. If $F_-^{-1}(u) = +\infty$, $u = 1$ and $F(F_-^{-1}(u)) = F(+\infty) = 1 = u$. If $F_-^{-1}(u) = -\infty$, $u = 0$ and $F(F_-^{-1}(u)) = F(-\infty) = 0 = u$, which concludes the proof of this point.

5. This point is easily deduced from the points 3 and 4.

6. If $F^{-1}(u) < F_-^{-1}(u)$, then

$$F\left(F^{-1}(u)\right) = \mathbb{P}\left(T \le F^{-1}(u)\right) \le \mathbb{P}\left(T < F_-^{-1}(u)\right) = F_-\left(F_-^{-1}(u)\right) \ .$$

The result then follows from point 2.

7. As $F_-(T) \le F(T)$, from point 3, one derives that:

$$\mathbb{1}_{\{F_-(T)=u\}} = \mathbb{1}_{\{F_-(T)=u\}} \mathbb{1}_{\left\{F^{-1}(u) \le T \le F_-^{-1}(u)\right\}} \ .$$

Since $\mathbb{P}\left(F^{-1}(u) < T < F_-^{-1}(u)\right) = F_-\left(F_-^{-1}(u)\right) - F\left(F^{-1}(u)\right) = 0$ (see point 2), one has:

$$\mathbb{1}_{\{F_-(T)=u\}} = \mathbb{1}_{\{F_-(T)=u\}} \mathbb{1}_{\left\{T \in \{F^{-1}(u), F_-^{-1}(u)\}\right\}} \quad \text{a.s.}$$

Furthermore, it is clear for every t in $\mathbb{R}$ (and even in the case $t = +/-\infty$) that

$$\mathbb{P}\left(T = t, F(T) = F_-(T)\right) = \mathbb{P}\left(T = t\right) \mathbb{1}_{\{F(t) = F_-(t)\}} = 0 \ . \tag{20}$$

So, applying (20) with $t = F^{-1}(u)$ and $t = F_-^{-1}(u)$, one obtains that

$$\mathbb{1}_{\{F_-(T)=u\}} = \mathbb{1}_{\{F_-(T)=u\}}\mathbb{1}_{\left\{T\in\{F^{-1}(u),F_-^{-1}(u)\}\right\}}\mathbb{1}_{\{F_-(T)<F(T)\}} \quad \text{a.s.}$$

Now let us prove that on the event

$$\Omega = \{F_-(T) = u,\ T \in \{F^{-1}(u), F_-^{-1}(u)\},\ F_-(T) < F(T)\}\ ,$$

T is necessarily equal to $F_-^{-1}(u)$.
Note that if $F^{-1}(u) = F_-^{-1}(u)$, there is nothing to prove, so the only case to consider is when $F^{-1}(u) < F_-^{-1}(u)$.
In this case, on the event Ω, if $T = F^{-1}(u)$, there exists $T' \in \mathcal{F}_u^-$ such that $T < T'$. Then, for such a T',

$$F_-(T) < F(T) \leq F_-(T') \leq u\ ,$$

which in fact contradicts $F_-(T) = u$. Therefore, on Ω, one actually necessarily has $T = F_-^{-1}(u)$. Hence,

$$\begin{aligned}\mathbb{1}_{\{F_-(T)=u\}}\mathbb{1}_{\left\{T\in\{F^{-1}(u),F_-^{-1}(u)\}\right\}}\mathbb{1}_{\{F_-(T)<F(T)\}} &\\ &= \mathbb{1}_{\{F_-(T)=u\}}\mathbb{1}_{\left\{T=F_-^{-1}(u)\right\}}\mathbb{1}_{\{F_-(T)<F(T)\}}\\ &= \mathbb{1}_{\{F_-(T)=u\}}\mathbb{1}_{\left\{T=F_-^{-1}(u)\right\}} \quad \text{a.s.}\ ,\end{aligned}$$

where the second equality comes from (20). This is enough to conclude that

$$\mathbb{1}_{\{F_-(T)=u\}} = \mathbb{1}_{\left\{T=F_-^{-1}(u)\right\}}\mathbb{1}_{\left\{F_-\left(F_-^{-1}(u)\right)=u\right\}} \quad \text{a.s.}$$

Magalie Fromont
Univ. Rennes 2
Place du Recteur H. Le Moal
35043 Rennes Cedex, France
E-mail: magalie.fromont@univ-rennes2.fr

Matthieu Lerasle
Univ. Nice Sophia Antipolis
LJAD
06100 Nice, France
E-mail: mlerasle@unice.fr

Patricia Reynaud-Bouret
Univ. Nice Sophia Antipolis
LJAD
06100 Nice, France
E-mail: reynaudb@unice.fr